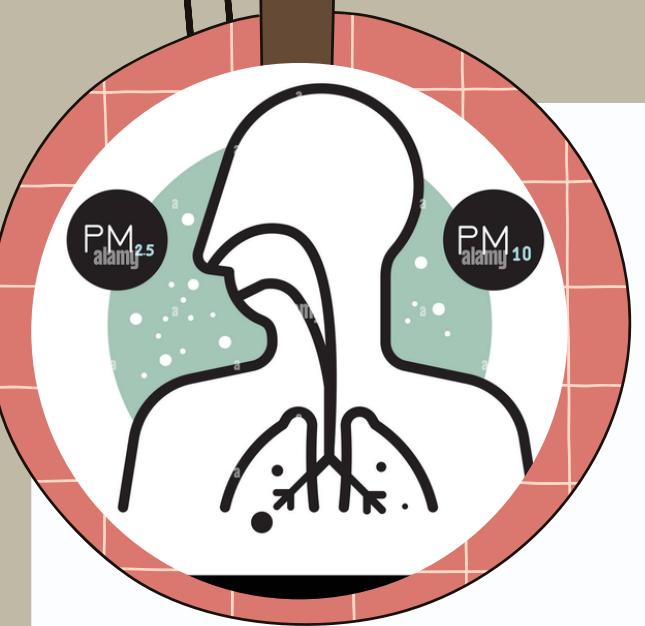




ESTRATEGIAS DE MODELADO PREDICTIVO DE CONCENTRACIÓN DE PARTÍCULAS FINAS A PARTIR DE VARIABLES SATELITALES

Autor: Rodríguez Núñez Martín

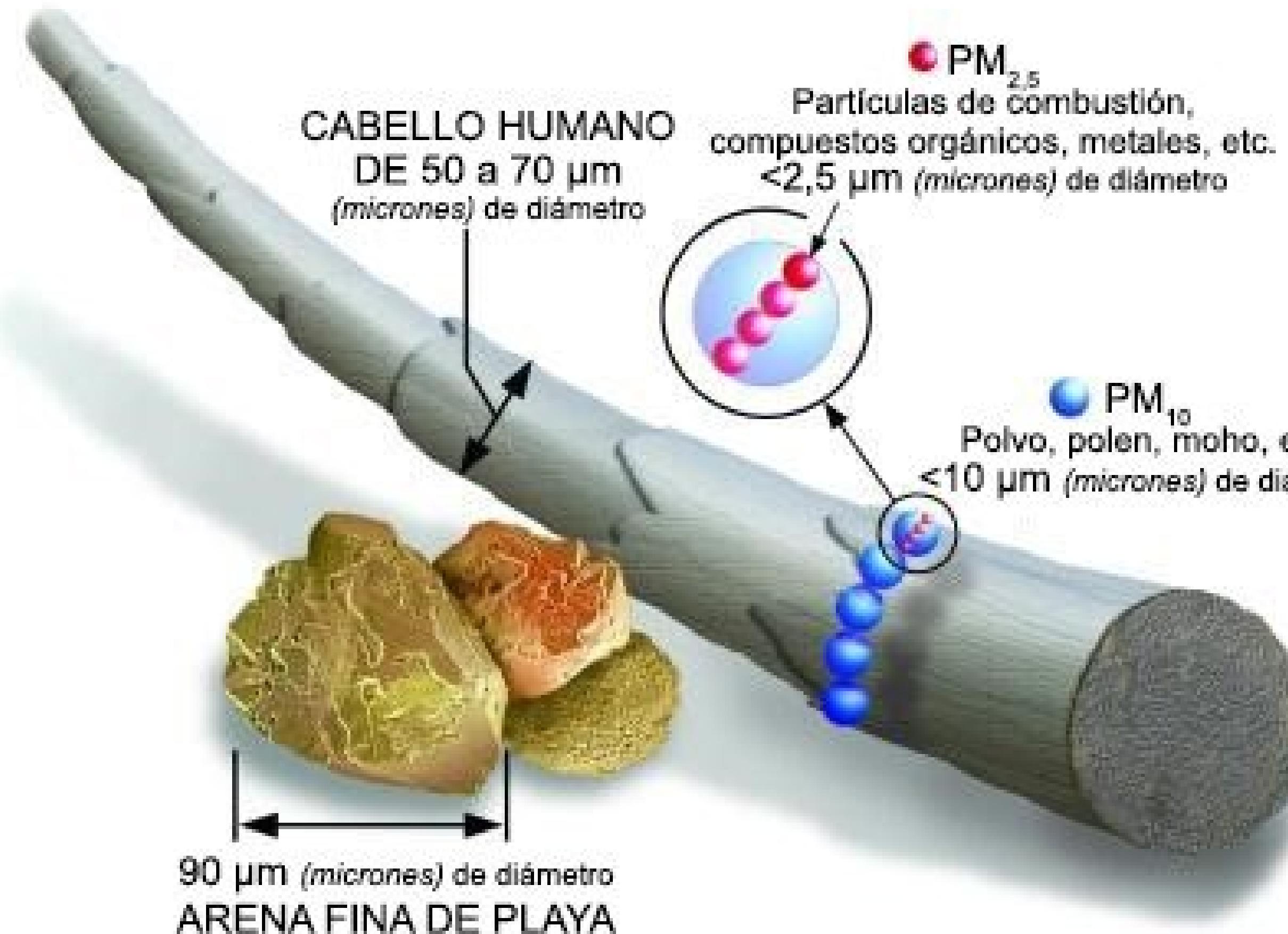
Directora: Mónica Balzarini



INTRODUCCIÓN

La Organización Mundial de la Salud (OMS) estima que nueve de cada diez personas respiran aire con altos niveles de contaminante y como resultado siete millones mueren cada año en el mundo por causas relacionadas.

El material particulado fino, partículas con un diámetro menor o igual a 2.5 micrómetros (PM_{2.5}), es considerado uno de los contaminantes atmosféricos de mayor nocividad dada su capacidad de penetrar la barrera pulmonar y pasar al torrente sanguíneo, causando enfermedades cardiovasculares, respiratorias y cáncer.



En el área de un corte transversal de un cabello humano
cabrían alrededor de 600 partículas de 2.5 µm de diámetro.

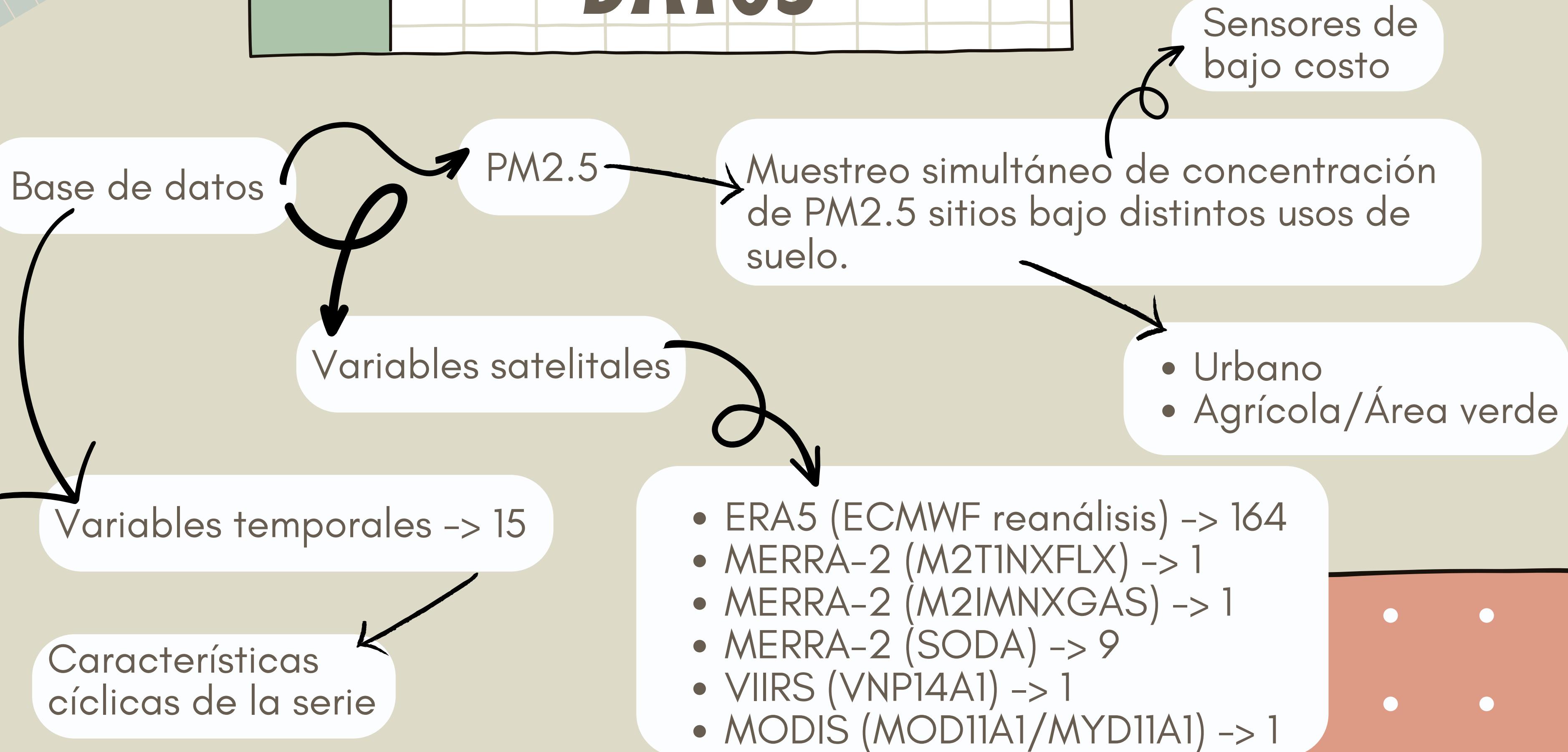
CONTEXTO

El monitoreo de la concentración de PM2.5 a través de sensores de referencia se ve limitado en países en vías de desarrollo dado a sus elevados costes de adquisición y mantenimiento, motivo por el cual se hace necesario recurrir a métodos alternativos para lograr monitorear su concentración y gestionar la calidad del aire.

OBJETIVO GENERAL

Explorar las capacidades predictivas de algoritmos de aprendizaje automático y profundo que emplean como entrada variables satelitales para pronosticar la concentración de PM2.5 en áreas con distinto uso de suelo en la provincia de Córdoba, Argentina.

DATOS



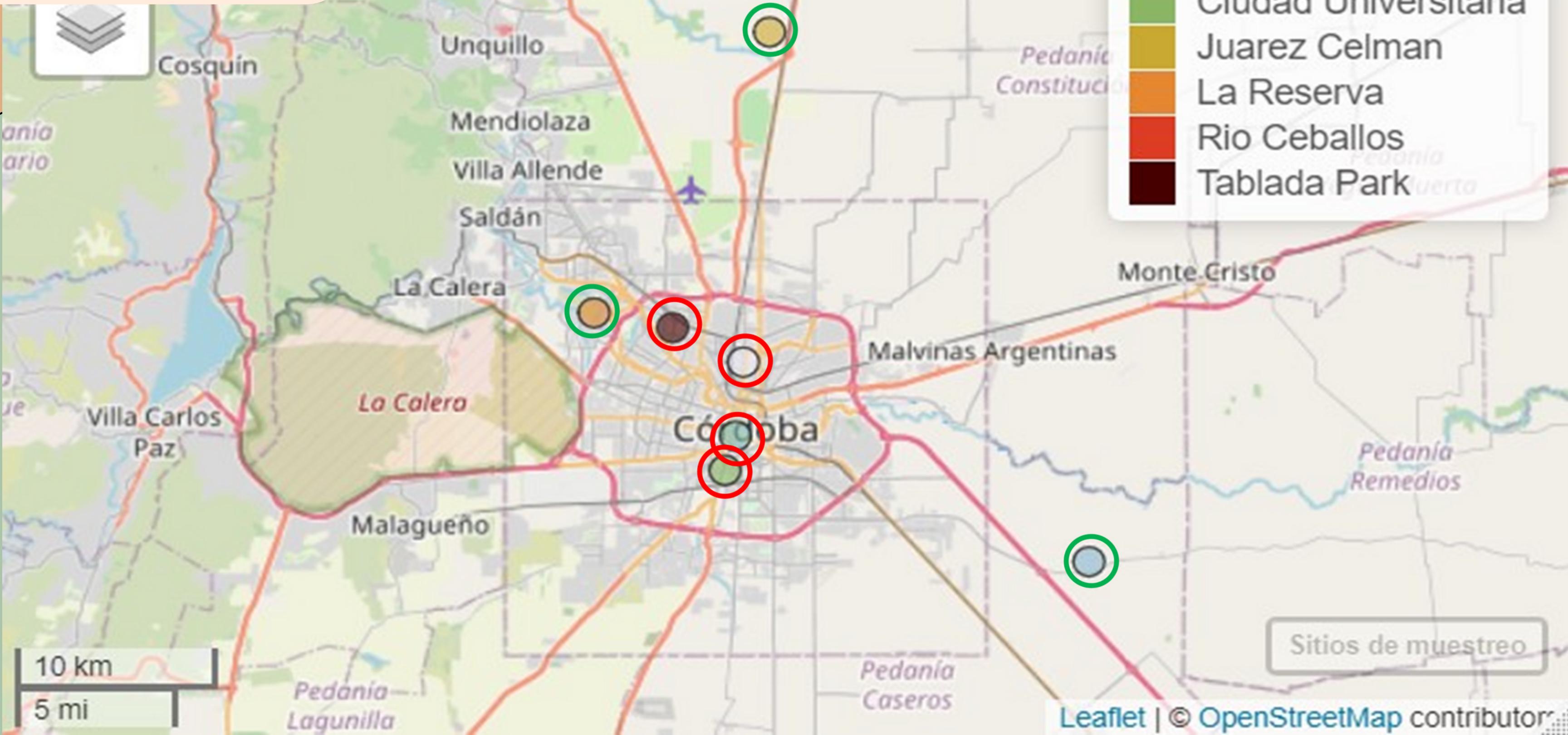


Urbano
Agrícola y área verde

- Mínimo 0.53 $\mu\text{g}/\text{m}^3$
- Máximo 150.28 $\mu\text{g}/\text{m}^3$
- Media 10.65 $\mu\text{g}/\text{m}^3$



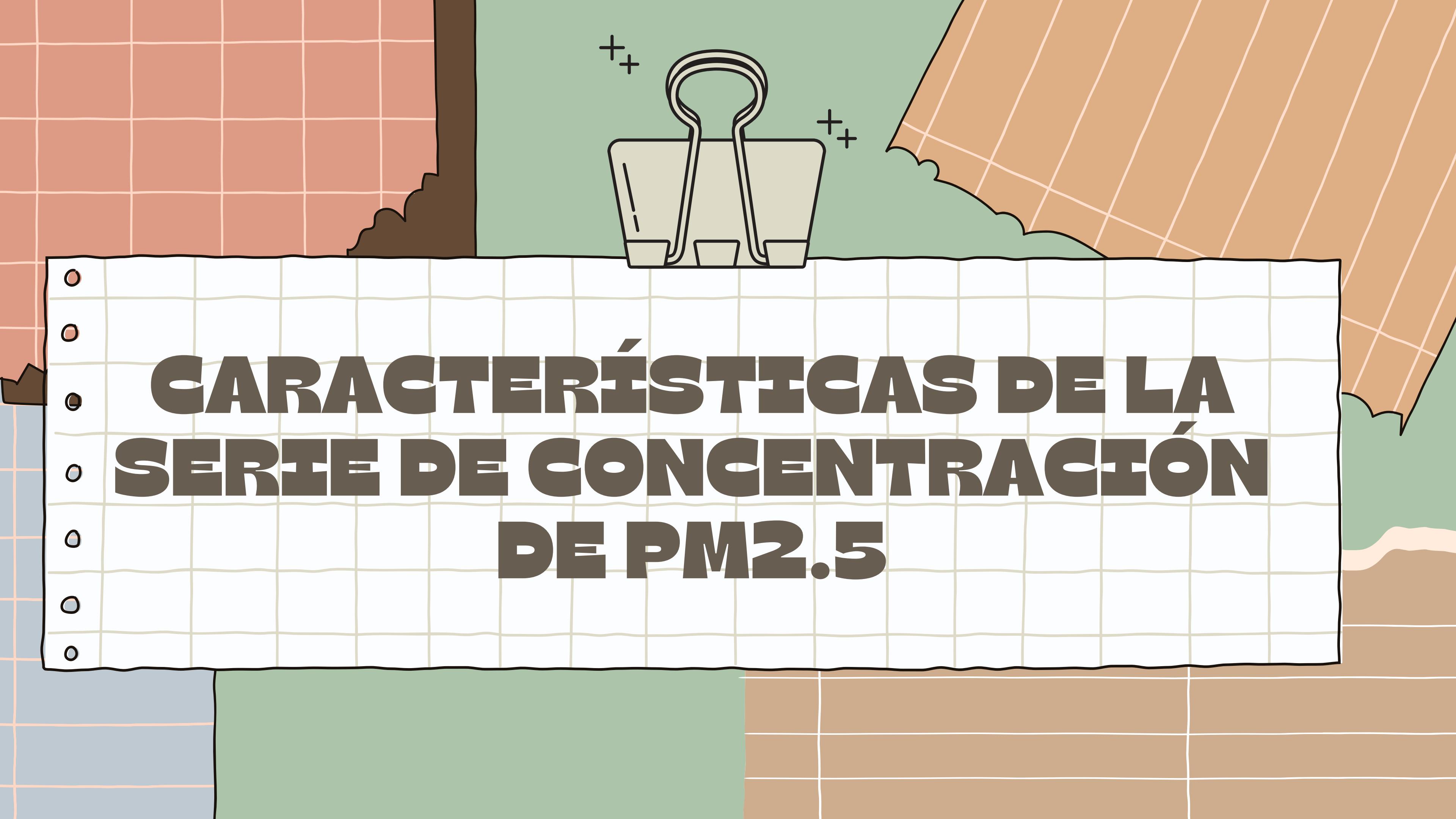
Cosquín



10 km

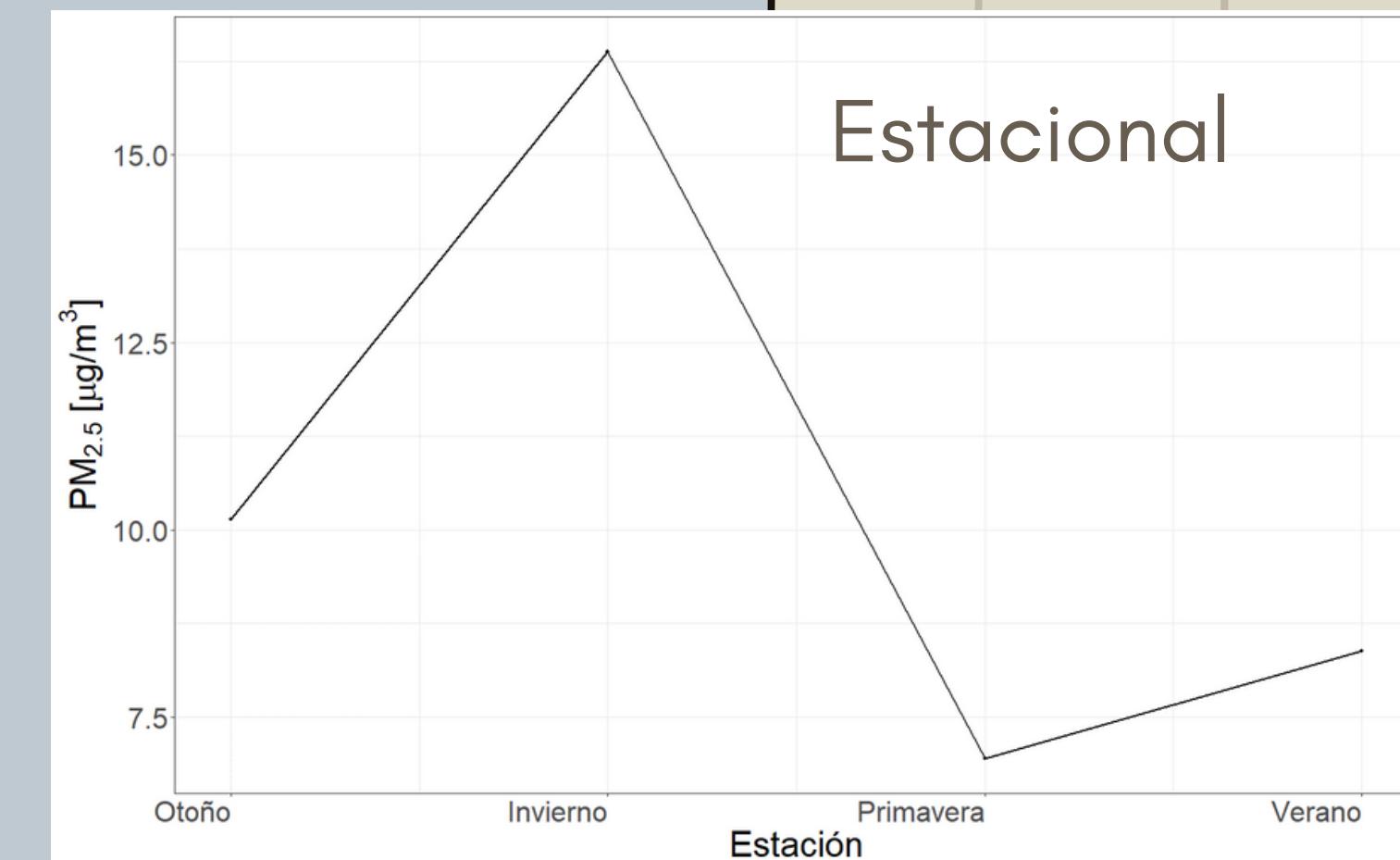
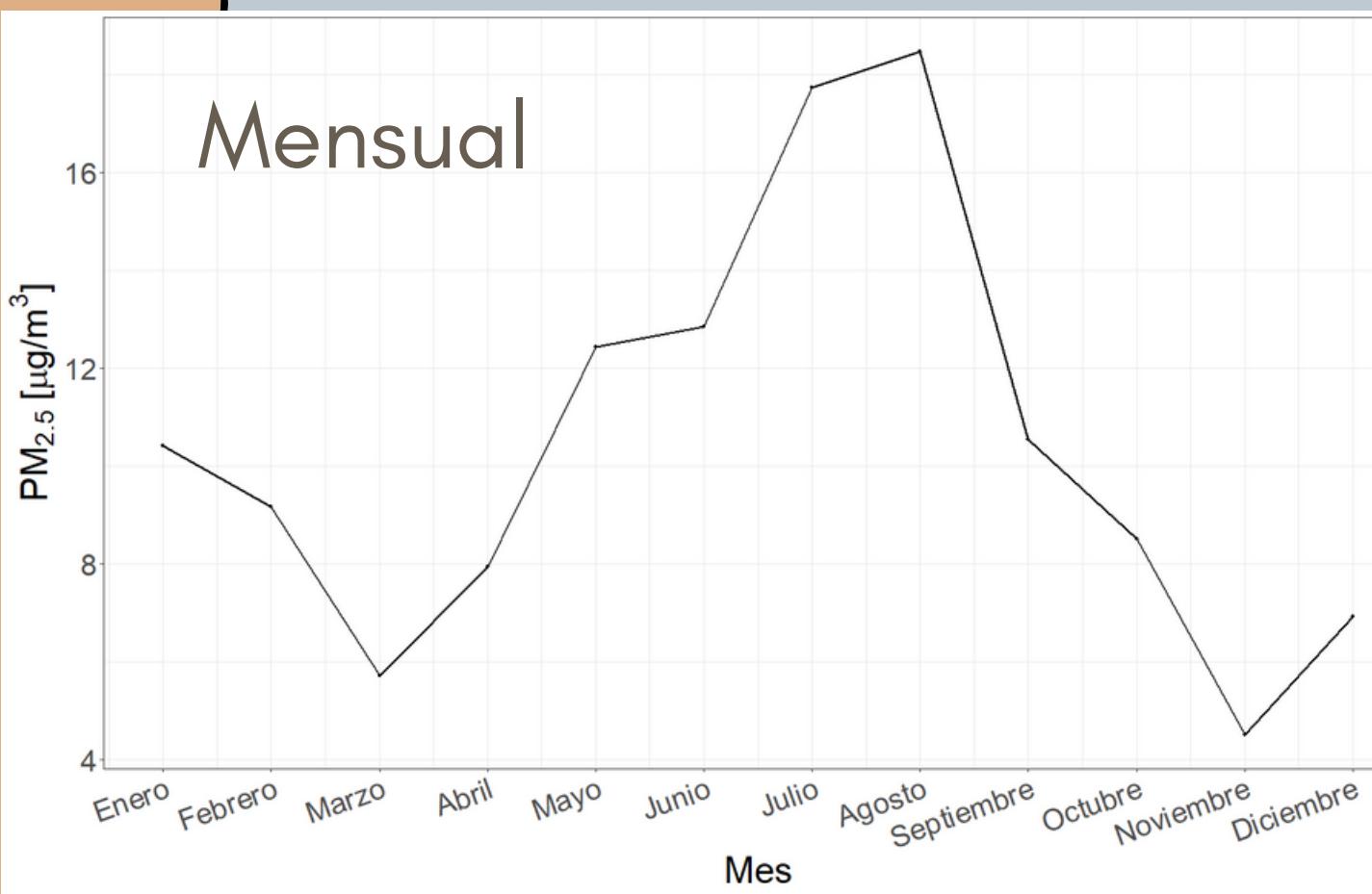
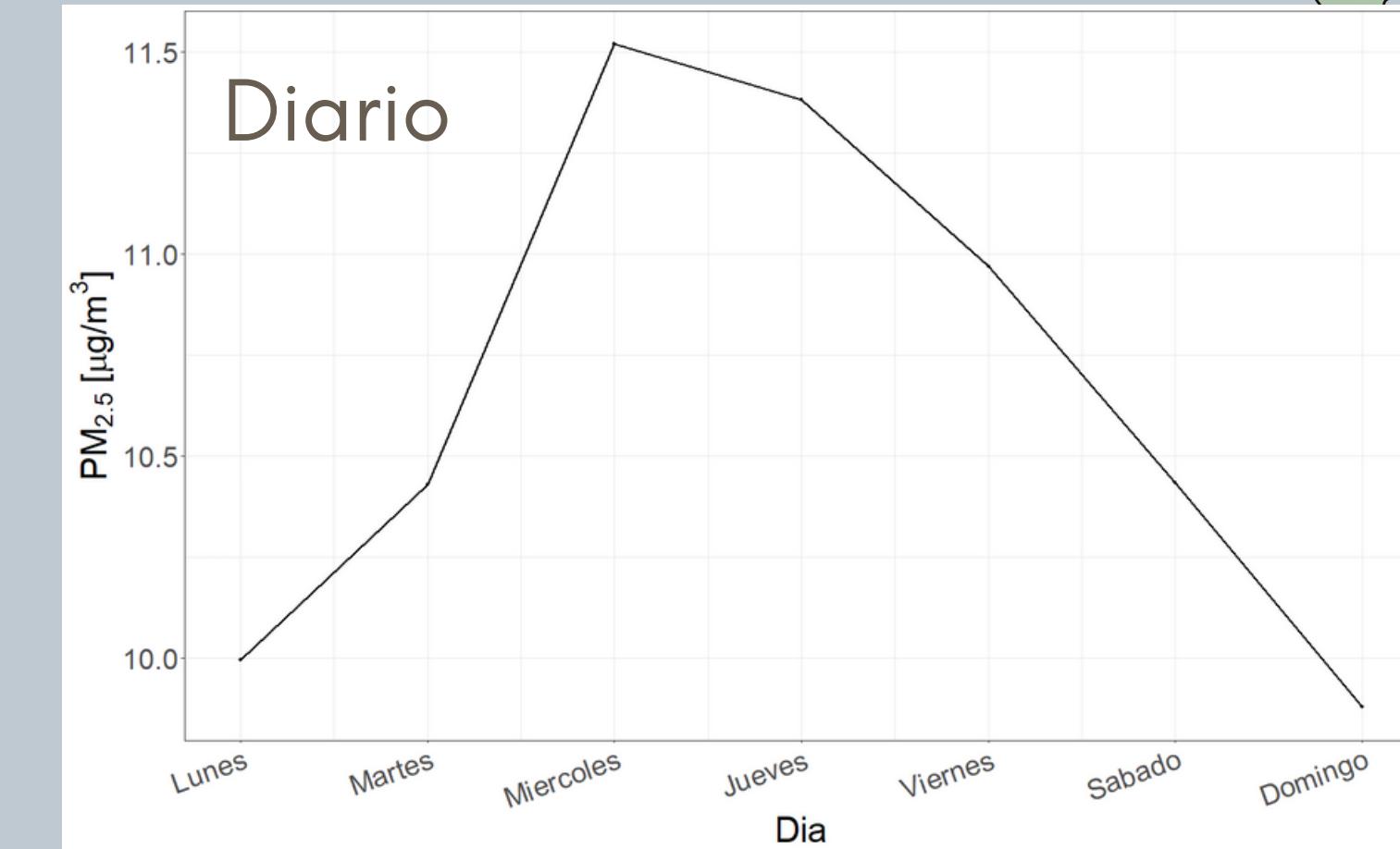
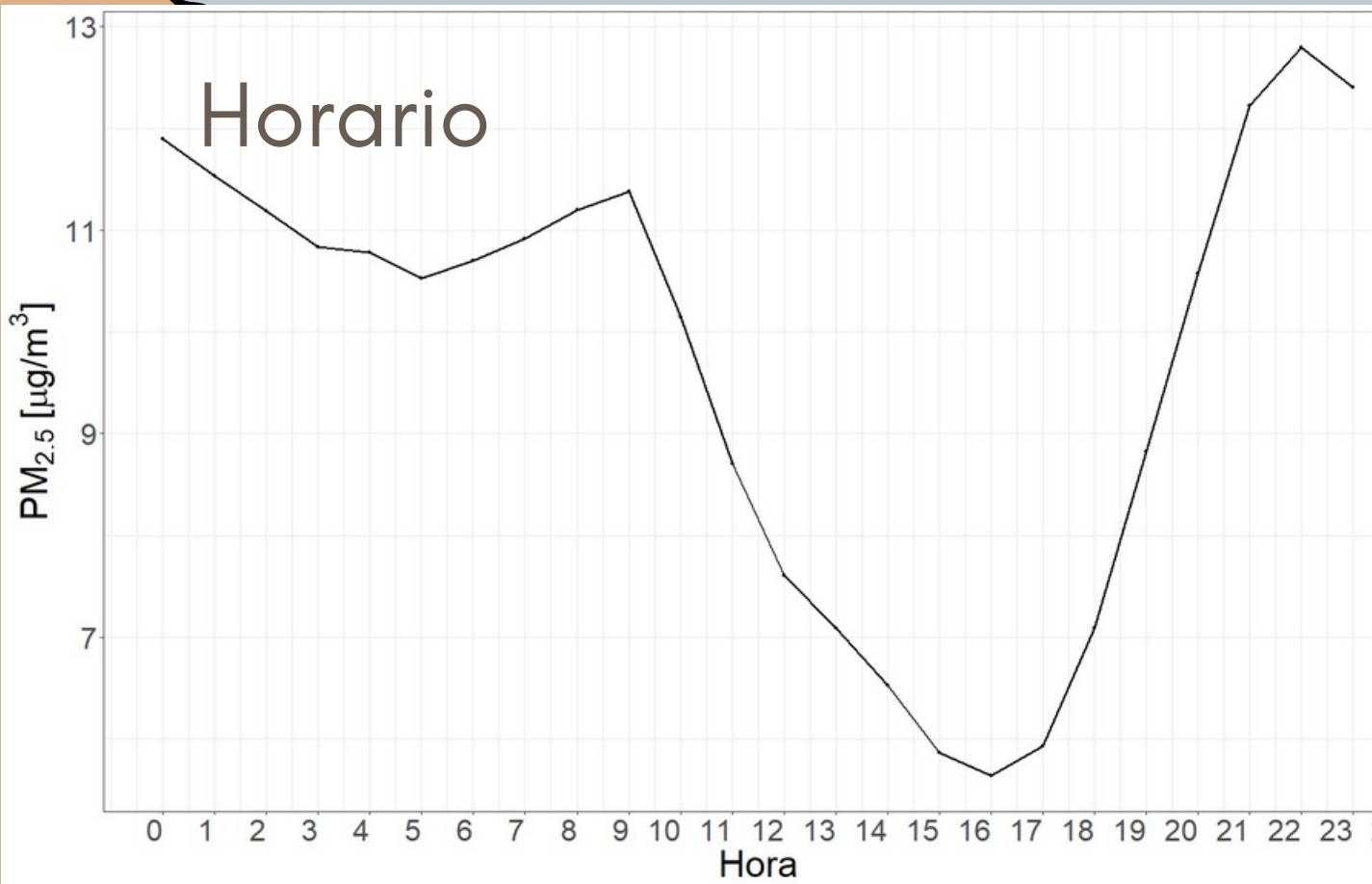
5 mi

Sitios de muestreo

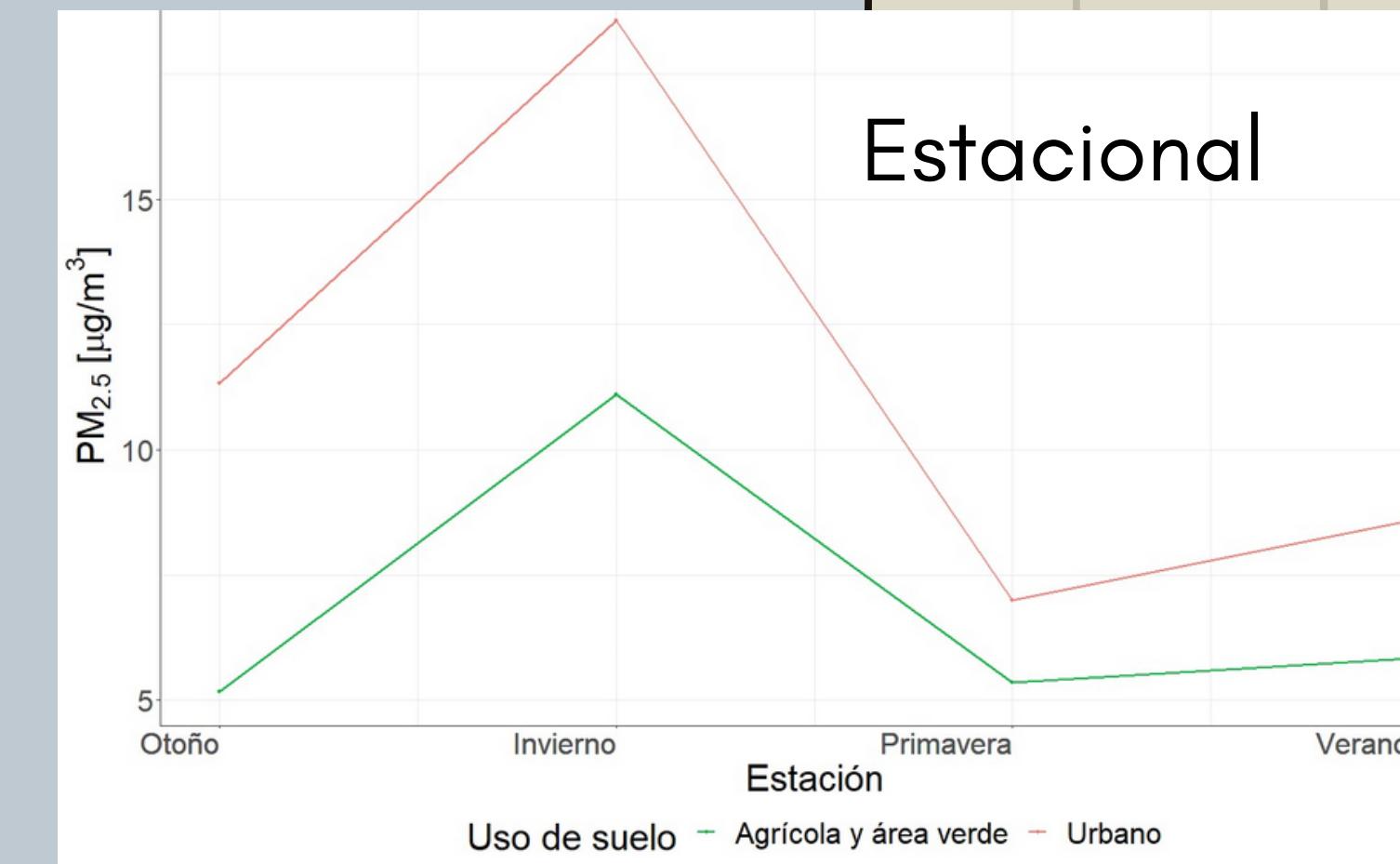
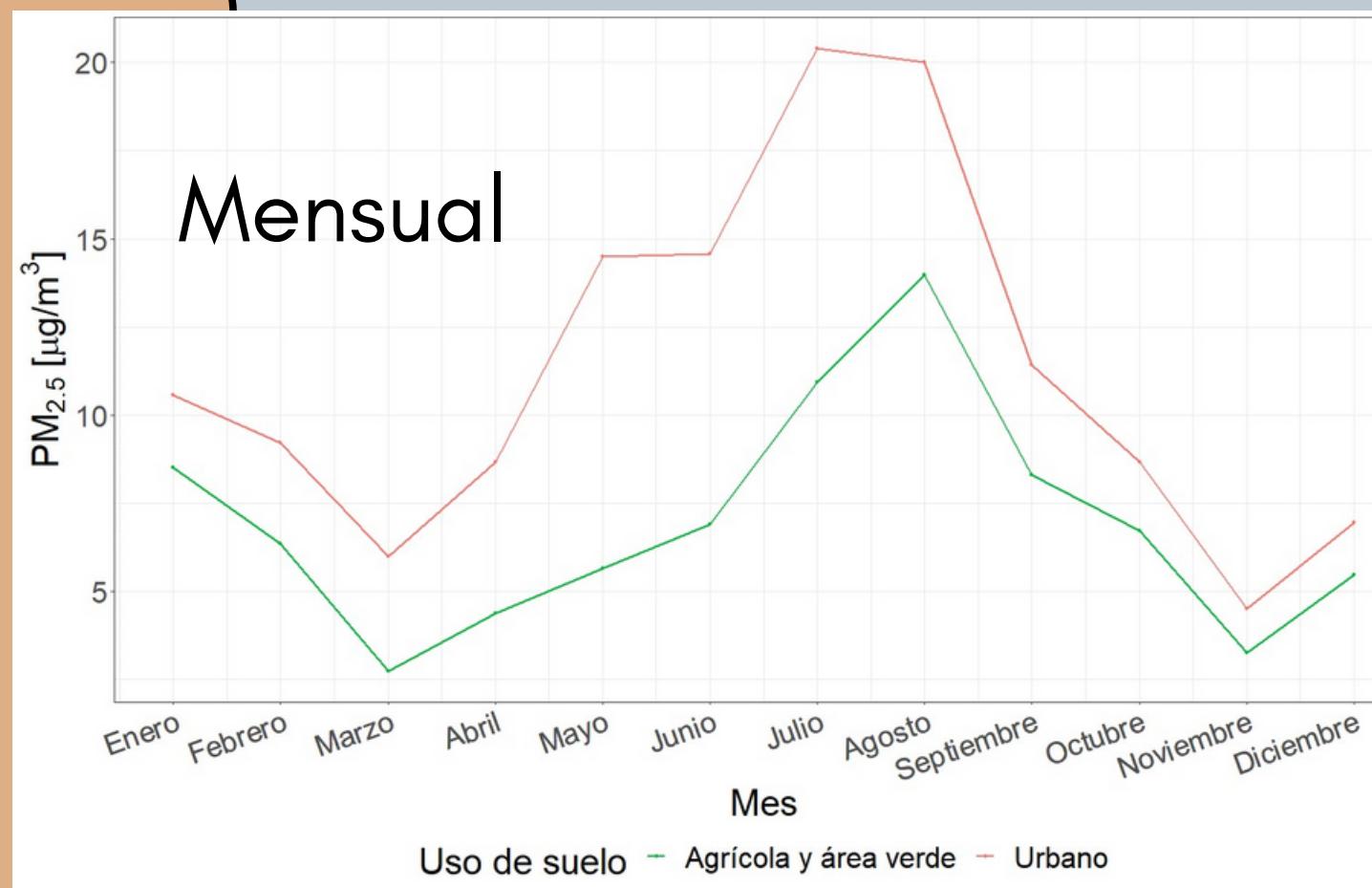
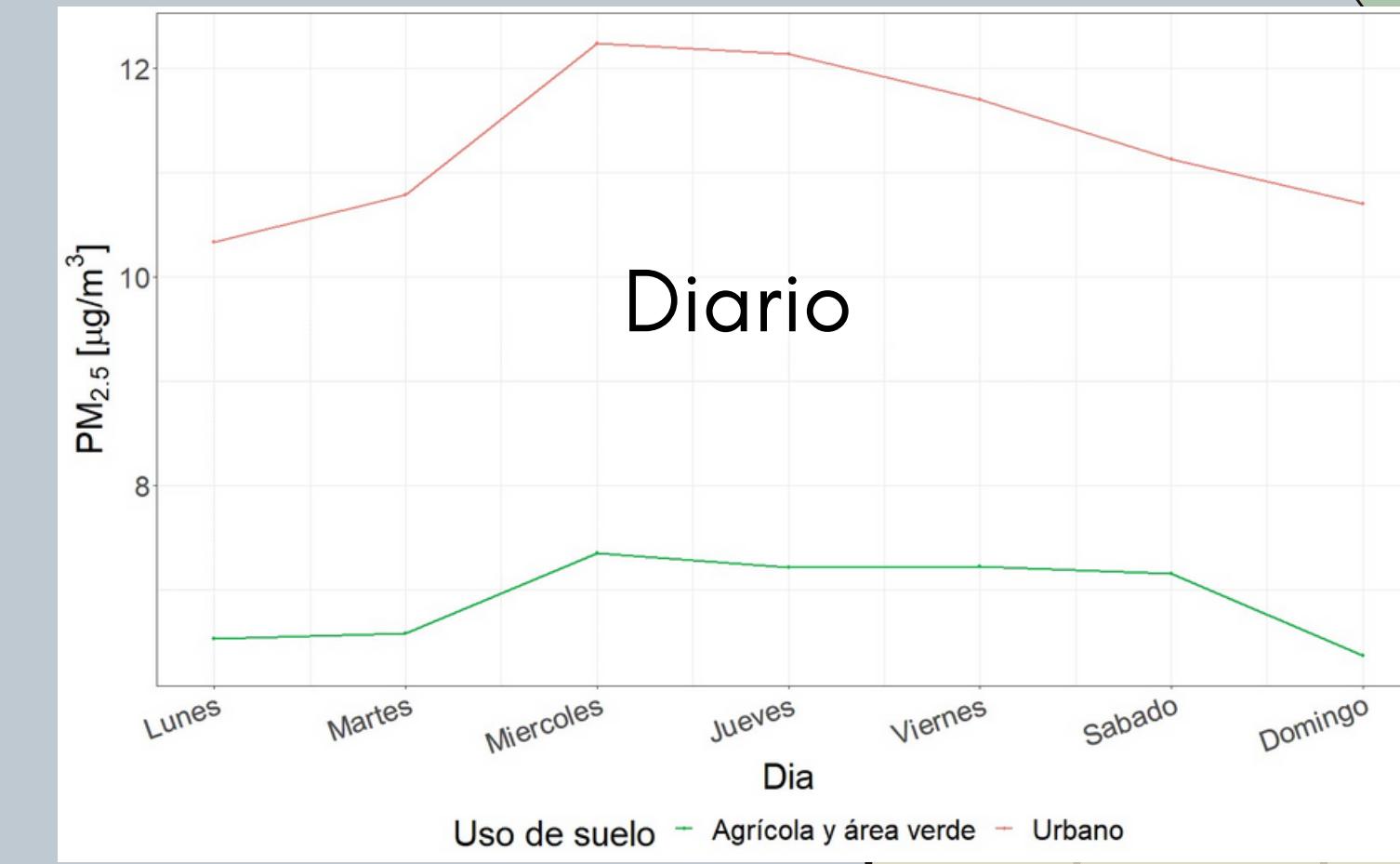
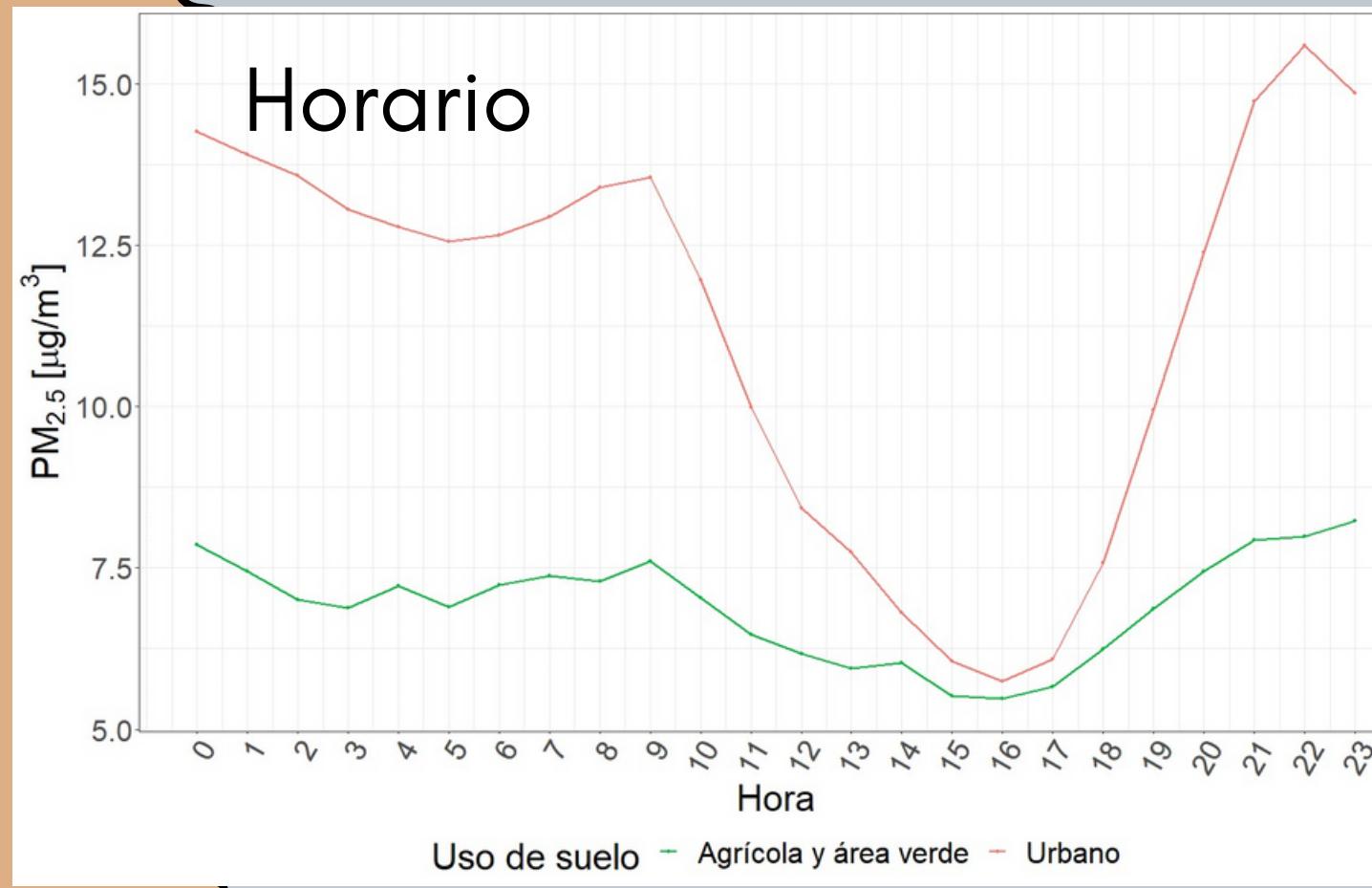


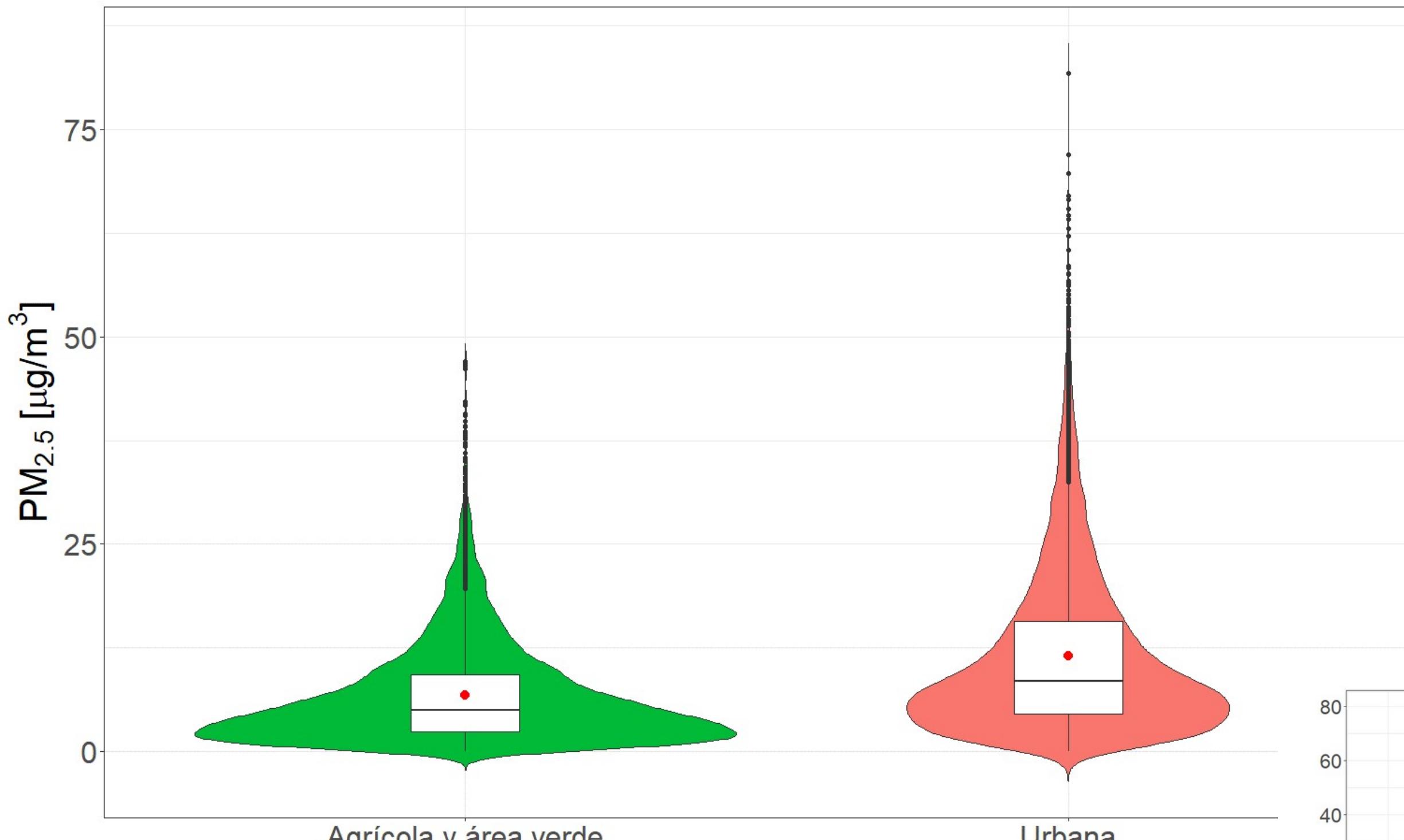
CARACTERÍSTICAS DE LA SERIE DE CONCENTRACIÓN DE PM_{2.5}

CARACTERÍSTICAS TEMPORALES



POR USO DE SUELO

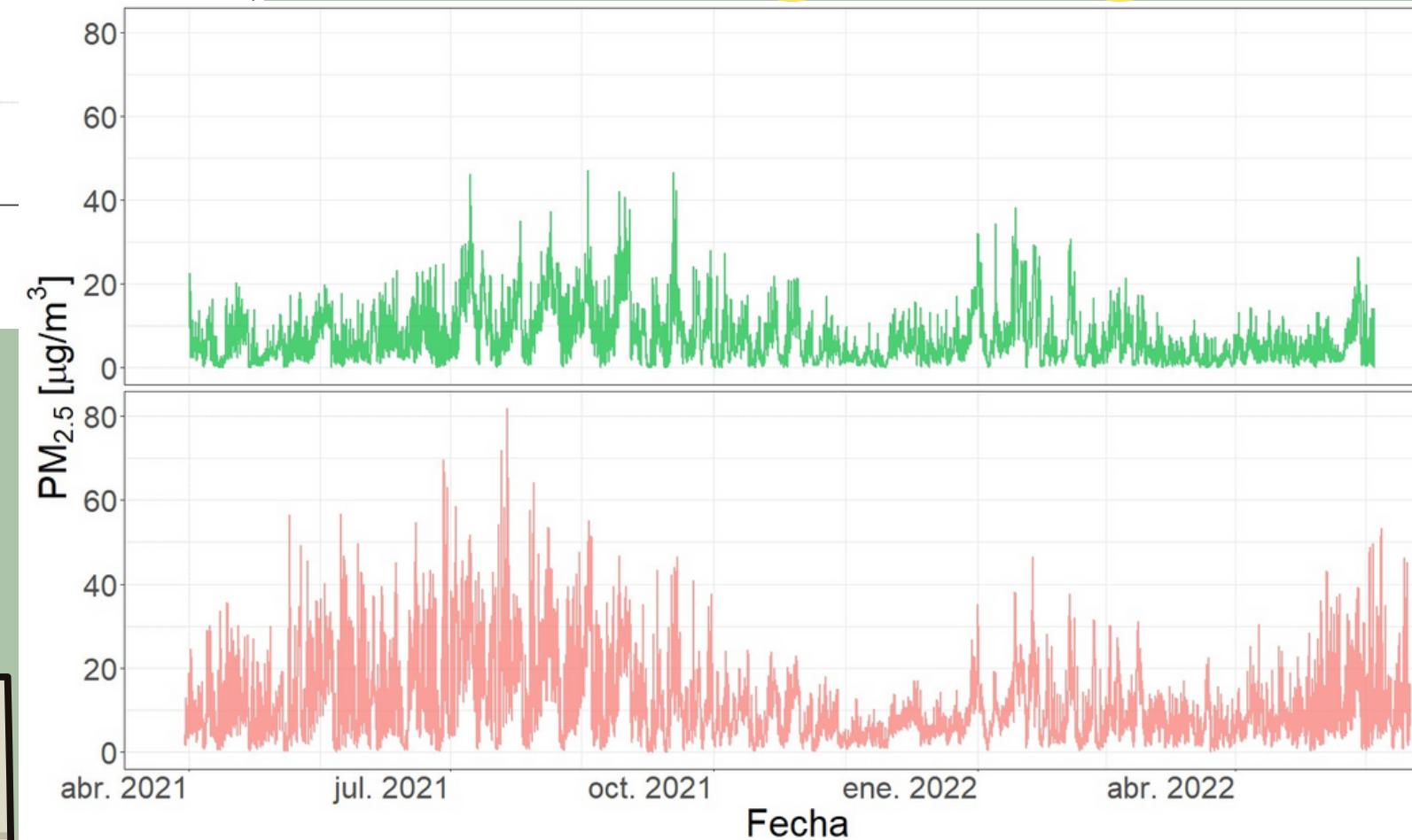




Percentil 90 agrícola y de área verde \approx percentil 75 urbana

RI urbano >> RI agrícola y de área verde

Medida	Urbano	Agrícola y área verde
Media	11.54	6.78
Desvió estándar	9.90	6.10
Máximo	81.80	47.02
Mínimo	0	0
Percentil 50	8.41	4.92
Percentil 75	15.72	9.30
Percentil 90	25.49	15.16
Percentil 95	32.11	19.57
RI	7.31	4.38



Características temporales

Patrón que describe la concentración de PM2.5

Entornos agrícolas y área verde

- Concentraciones menor magnitud.
- Concentración de datos en torno de su media.
- Menores fluctuaciones.

Variación en la magnitud de las concentraciones

Independiente del uso de suelo

Serie temporal de menor variabilidad

Uso de suelo

Características de la microescala

Influencia de las condiciones meteorológicas

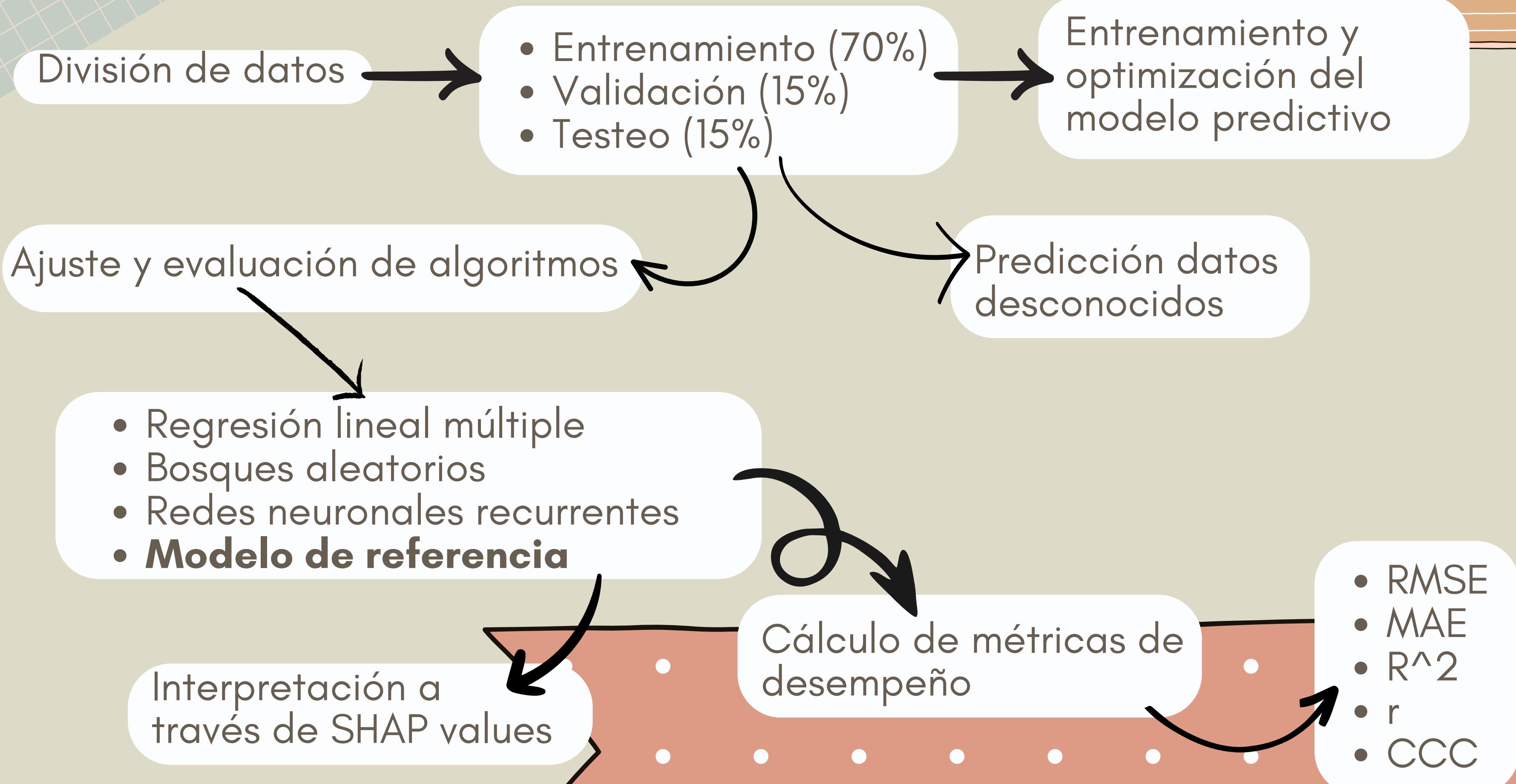
- Fuente de emisión.
- Topología
- ...

Urbana

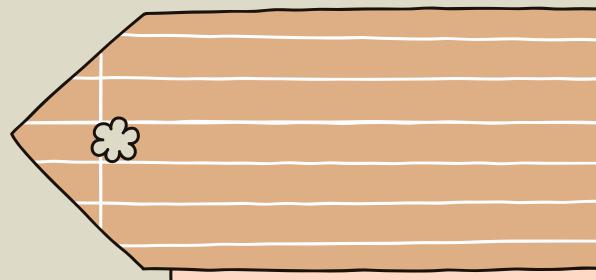
MODELADO PREDICTIVO



MODELADO PREDICTIVO



Evaluar consistencia en los resultados obtenidos



Distintas bases de datos

A lo largo de cada uno de los sitios en ese uso de suelo

- Base de datos original
- Algoritmo LASSO -> BD 1
- Featurewiz -> BD 2

Coeficiente de regularización

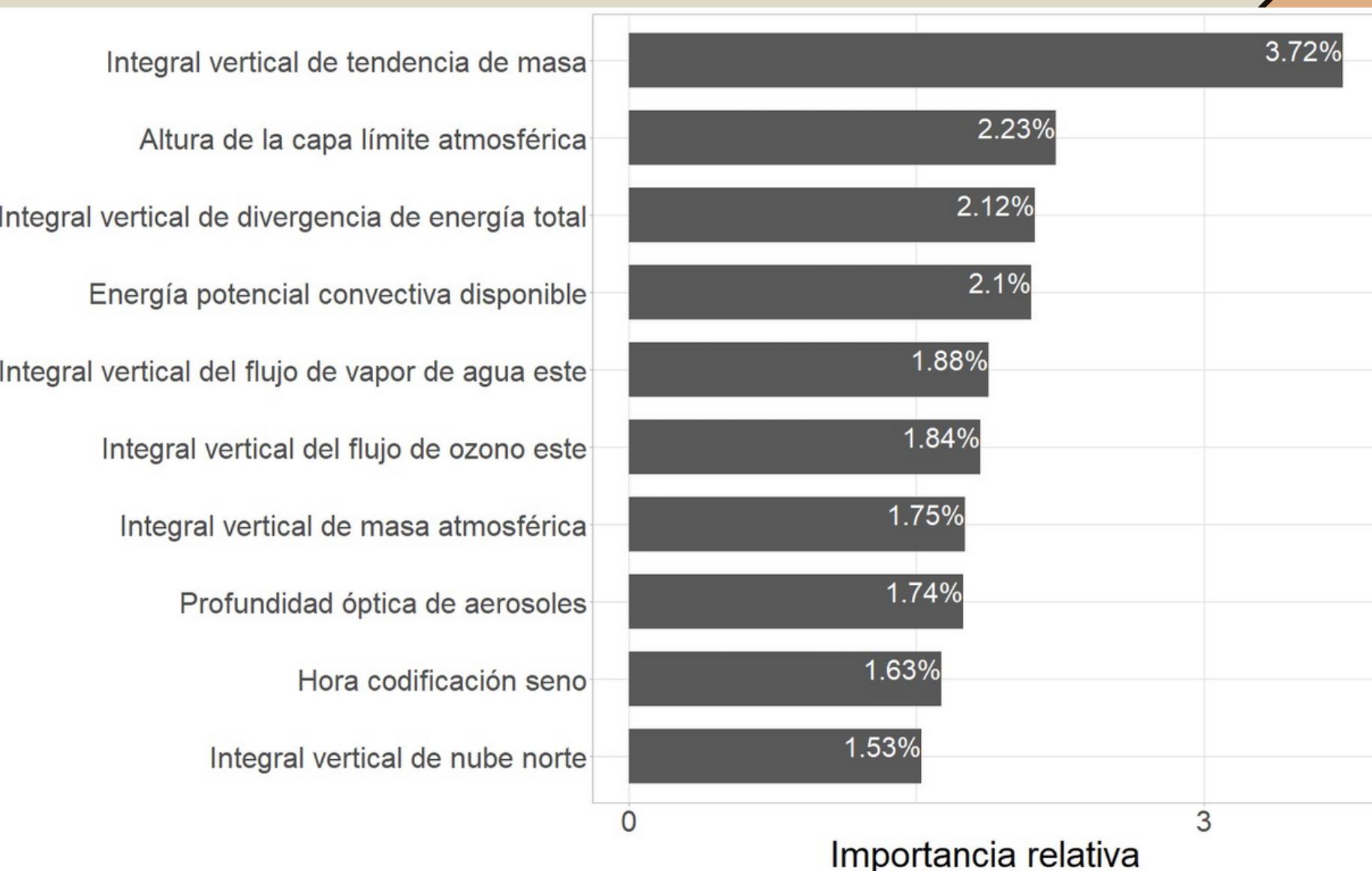
Combina el algoritmo de SULOV (Searching for Uncorrelated List of Variables), basado en MRMR (Mínima Redundancia Máxima Relevancia) y una eliminación recursiva de características basada en XGBoost.

URBANO

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Redes neuronales recurrentes	1.657	1.174	0.972	0.987	0.986
Bosques Aleatorios	6.823	4.919	0.529	0.754	0.639
Regresión lineal múltiple	7.840	5.788	0.378	0.615	0.548
<i>Modelo de referencia</i>	3.740	2.316	0.859	0.929	0.929

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Redes neuronales recurrentes	3.747	2.472	0.857	0.927	0.924
Bosques Aleatorios	7.135	5.168	0.482	0.715	0.601
Regresión lineal múltiple	7.929	5.888	0.360	0.601	0.540
<i>Modelo de referencia</i>	3.740	2.326	0.858	0.930	0.930

Resultados similares para las distintas versiones de bases de datos y sitios.

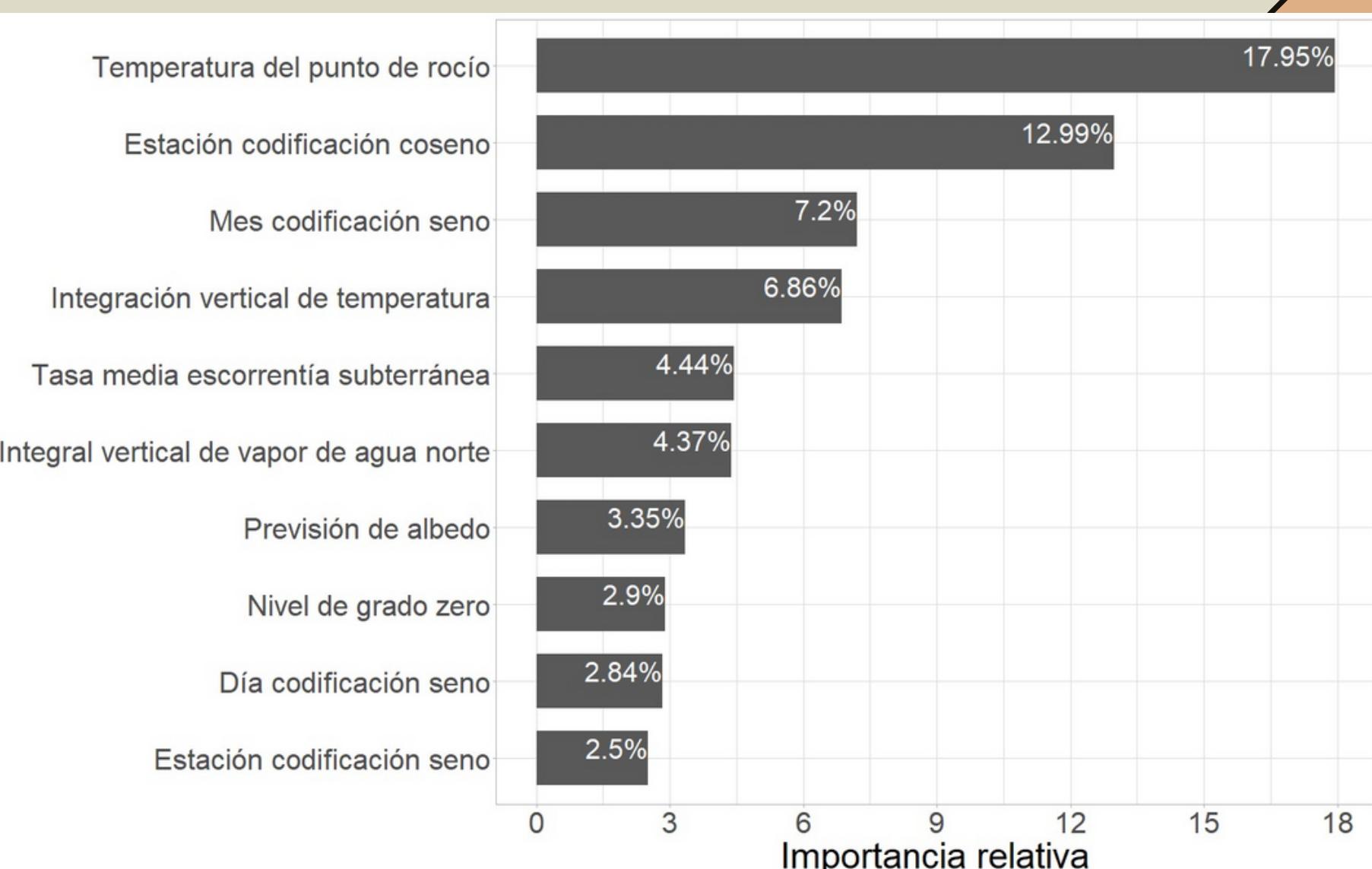


AGRÍCOLA / ÁREA VERDE

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Bosques Aleatorios	0.975	0.625	0.975	0.990	0.986
Redes neuronales recurrentes	1.530	1.051	0.937	0.970	0.967
Regresión lineal múltiple	5.250	3.879	0.273	0.523	0.429
<i>Modelo de referencia</i>	2.813	1.569	0.792	0.896	0.896

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Bosques Aleatorios	2.550	1.650	0.827	0.912	0.899
Redes neuronales recurrentes	2.784	1.796	0.794	0.893	0.888
Regresión lineal múltiple	5.241	3.900	0.272	0.523	0.430
<i>Modelo de referencia</i>	2.718	1.554	0.804	0.903	0.903

Resultados similares para las distintas versiones de bases de datos y sitios.



USO DE SUELO

Redes
neuronales
recurrentes

Urbano

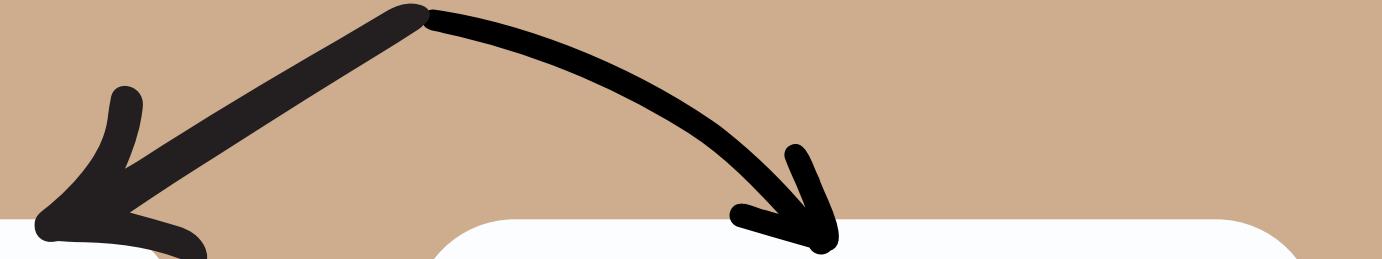
Agrícola/
Área verde

Bosques
aleatorios
(random forest)

Algoritmo predictivo
de mejor ajuste

Características de las series
temporales en cada uso de suelo

Características de aprendizaje
de cada uno de los algoritmos



Similitud métricas

Modelo de referencia

Modelos predictivos desarrollados

Consistencia
de los
resultados

Series temporales
de menor
variabilidad

Bosques
aleatorios

Distintos enfoques de bases
de datos y sitios de muestreo

Series temporales
de mayor
variabilidad

Potencial para alertar
a la población ante
condiciones adversas
de calidad de aire.

Redes
neuronales
recurrentes



FCA
Facultad de Ciencias
Agropecuarias



Universidad
Nacional
de Córdoba



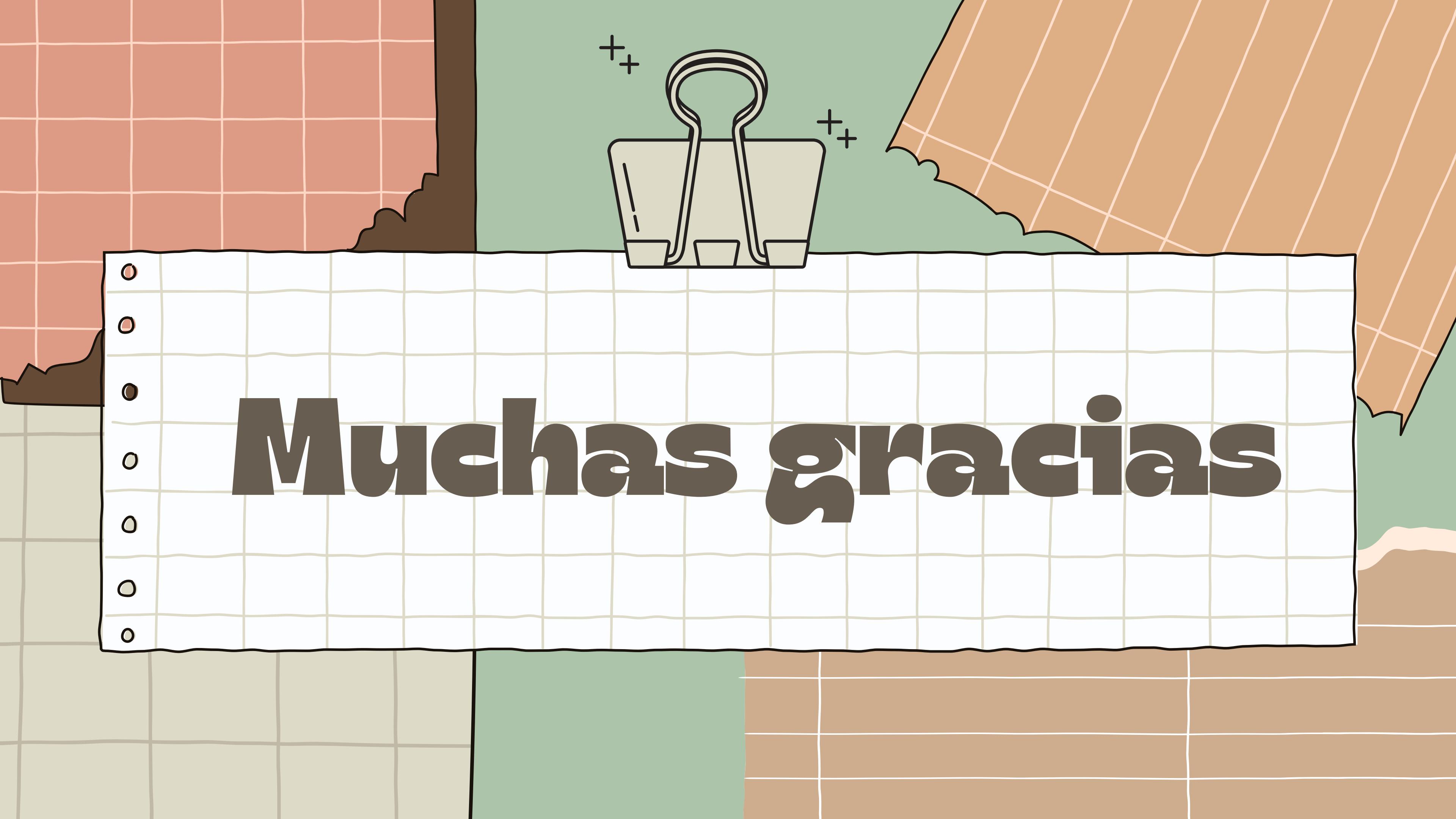
FCEFyN

CONICET



AGRADECIMIENTOS





Muchas gracias